

Published in final edited form as:

Phys Rev E Stat Nonlin Soft Matter Phys. 2012 September ; 86(3 0 2): 036104.

Taxonomies of networks from community structure

Jukka-Pekka Onnela^{#1,2,3,4}, Daniel J. Fenn^{#5,4}, Stephen Reid³, Mason A. Porter^{6,4}, Peter J. Mucha⁷, Mark D. Fricker^{8,4}, and Nick S. Jones^{3,9,4}

¹Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA

²Department of Health Care Policy, Harvard Medical School, Boston, MA 02115, USA

³Department of Physics, University of Oxford, Oxford OX1 3PU, UK

⁴CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK

⁵Mathematical and Computational Finance Group, University of Oxford, Oxford OX1 3LB, UK

⁶Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, OX1 3LB, UK

⁷Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics and Institute for Advanced Materials, Nanoscience & Technology, University of North Carolina, Chapel Hill, NC 27599, USA

⁸Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

⁹Oxford Centre for Integrative Systems Biology, Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK

These authors contributed equally to this work.

Abstract

The study of networks has become a substantial interdisciplinary endeavor that encompasses myriad disciplines in the natural, social, and information sciences. Here we introduce a framework for constructing taxonomies of networks based on their structural similarities. These networks can arise from any of numerous sources: they can be empirical or synthetic, they can arise from multiple realizations of a single process (either empirical or synthetic), they can represent entirely different systems in different disciplines, etc. Because mesoscopic properties of networks are hypothesized to be important for network function, we base our comparisons on summaries of network community structures. Although we use a specific method for uncovering network communities, much of the introduced framework is independent of that choice. After introducing the framework, we apply it to construct a taxonomy for 746 networks and demonstrate that our approach usefully identifies similar networks. We also construct taxonomies within individual categories of networks, and we thereby expose nontrivial structure. For example, we create taxonomies for similarity networks constructed from both political voting data and financial data. We also construct network taxonomies to compare the social structures of 100 Facebook networks and the growth structures produced by different types of fungi.

Keywords

networks; clustering; community structure

I. INTRODUCTION

Although there is a long tradition of scholarship on networks, the last two decades have witnessed substantial advances in network science due to developments in physics, mathematics, computer science, sociology, and numerous other disciplines [1, 2]. Given that the questions asked by researchers in different fields can be surprisingly similar, it would be useful to be able to highlight similarities in network structures across disciplines in a systematic way. One way to approach this is to formulate a suitable means of comparing networks and to use this means to develop taxonomies of networks. Such taxonomies have the potential to facilitate the identification of problems from different disciplines that might be approached similarly in terms of both empirical analyses and theoretical modeling. For example, if a biological network depicting covariation of neural activity in different regions of the brain is demonstrated to be structurally similar to a financial network representing correlations of stock returns, then certain types of edge thresholding methods or structural null models might be applicable to both situations.

From a historical perspective, classification of objects has often been central to the progress of science, as demonstrated by the periodic table of elements in chemistry and phylogenetic trees of organisms in biology [3]. It is plausible that an organization of networks has the potential to shed light on mechanisms for generating networks, reveal how an unknown network should be treated once one has discerned its position in a taxonomy, or help identify a network family's anomalous members. Further potential applications of network taxonomies include un-supervised study of multiple realizations of a given model process (e.g., characterizing the similarities and differences of many different networks drawn from the Erdős-Rényi random graph model using the same parameter values), examination of multiple empirical networks with known similar origins or generative processes, and the detection of anomalous changes in temporally ordered series of networks. In this paper, we develop a framework for the creation of network taxonomies [4]. In so doing, we develop the requisite diagnostic tools and discuss several case studies that suggest how our methodology can help illuminate relationships both between and within families of networks.

In aiming to construct taxonomies of networks, one has to consider the scales at which one wants to compare differences in network structures. Much research has focused on extremes—either microscopic (e.g., node degree) or macroscopic (e.g., mean geodesic distance) properties—and numerous researchers have, for example, reported that many empirical networks possess heavy-tailed degree distributions or the small-world property [1, 5]. Given the ubiquity of such findings, it is clear that more nuanced approaches are needed to make useful comparisons between networks. Indeed, interpretations of microscopic and macroscopic approaches often implicitly assume that networks are homogeneous and ignore “mesoscopic” structures in networks. To overcome some of these limitations, earlier work

has focused on the statistics of small, *a priori* specified modules called “motifs” [32, 37], role-to-role connectivity profiles of nodes [8], the isolation of statistically significant structures called “backbones” [9], interrelations of network modules [10], examination of the number of nodes located within “shells” [11], and the self-similarity of networks as characterized by fractal exponents [12]. The taxonomic framework that we develop in the present paper builds on the idea of examining network modules by computing community structures [13, 14], as was also done in the work of [15], and we subsequently compare signatures derived from community structure across networks. Importantly, although we use a specific method to uncover network communities, much of the introduced framework is independent of that choice. Consequently, our comparative framework can accommodate a large variety of community detection schemes.

The remainder of this paper is organized as follows. First, we discuss the detection of communities in networks in order to find coherent groups of nodes that are densely connected to each other. We then introduce *mesoscopic response functions* (MRFs), which allow us to probe how the community structure of a network changes as a function of a resolution parameter that determines network scales of interest. We then illustrate MRFs using several examples of networks and compare the MRFs for several well-known generative models of networks. We use MRFs to develop a means to measure distance between a pair of networks, and use this comparative measure to cluster networks and thereby develop taxonomies. Using 746 networks from numerous different fields, we construct a taxonomy of these networks. We then construct taxonomies of networks within fields using several case studies: voting in the United States Senate, voting in the United Nations General Assembly, Facebook networks at US universities, fungal networks, and networks of stock returns in the New York Stock Exchange. In each example, we expose structure that is either illuminating or can be checked against information from an external source (e.g., previously published investigations). This suggests that our method for comparing networks is capturing important similarities and differences. We conclude with a brief summary and discussion of our results. In addition, we provide further details in the Appendices and Supplemental Material. Among other topics, we examine the robustness of the obtained taxonomies, address some computational issues, tabulate some of the basic properties of the networks that we investigated, and provide references for the network data sources used in this study.

II. MULTI-RESOLUTION COMMUNITY DETECTION

Our approach is based on network *community structure* [13, 14]. A community consists of a set of nodes for which there are more edges (or, in the case of weighted networks, a greater total edge weight) connecting the nodes in the set than what would be expected by chance. The algorithmic detection of communities is a particularly active area of network science, in part because communities are thought to be related to functional units in many networks and in part because they can strongly influence dynamical processes that operate on networks [13, 14].

In this paper, we detect communities using the multi-resolution Potts method [13, 14, 16], a generalization of modularity optimization [9, 13, 14, 16, 18, 19]. (Modularity optimization is

perhaps the most popular approach for detecting communities.) Given a network adjacency matrix A_{ij} , we find communities by minimizing the Hamiltonian of the infinite-range N -state Potts spin glass

$$\begin{aligned}\mathcal{H}(\lambda) &= -\sum_{i \neq j} J_{ij}(\lambda) \delta(C_i, C_j) \\ &= -\sum_{i \neq j} (A_{ij} - \lambda P_{ij}) \delta(C_i, C_j),\end{aligned}\quad (1)$$

where C_i indicates the community (state) of node (spin) i , λ is a resolution parameter, and $\mathbf{J}(\lambda)$ is the coupling matrix with entries $J_{ij}(\lambda)$ representing the interaction strength between node i and node j in the Potts Hamiltonian. We use the (undirected-network) null model $P_{ij} = k_i k_j / (2m)$, where k_i denotes the strength (total edge weight) of node i and m is the total edge weight in the network [9]. By tuning the resolution parameter λ , we can detect communities at multiple scales of a network. Our particular choice of J_{ij} implies that we are optimizing modularity (with the addition of the resolution parameter) [13, 14].

To compare networks, we create profiles of summary statistics that characterize the community structure of each network at different mesoscopic scales. We also study a wide variety of networks that contain different numbers of nodes and edges. (We enumerate the networks that we consider in Table II of the Supplemental Material.) To ensure that we can compare the profiles for different networks, we sweep the resolution parameter λ from a minimum value Λ_{\min} to a maximum value Λ_{\max} (discussed in detail below). We define these quantities separately for each network such that the number of communities η into which the network is partitioned is 1 at Λ_{\min} and is equal to the total number of nodes N at Λ_{\max} . In other words, one can think of λ as a parameter that controls the fragmentation of a network into communities.

To find the minimum and maximum resolution-parameter values, consider the interactions in Eq. (1). An interaction is called *ferromagnetic* when $J_{ij} > 0$ and *antiferromagnetic* when $J_{ij} < 0$. For each pair of nodes i and j , we find the resolution $\lambda = \Lambda_{ij}$ at which the interaction J_{ij} is neutral (i.e., $J_{ij}(\Lambda_{ij}) = 0$), leading to $\Lambda_{ij} = A_{ij}/P_{ij}$. We thereby identify two special resolutions:

$$\Lambda_{\min} = \max_{ij} \{\Lambda_{ij} | \eta(\lambda) = 1\}, \quad (2)$$

$$\Lambda_{\max} = \max_{ij} \{\Lambda_{ij}\} + \epsilon, \quad (3)$$

where $\epsilon > 0$ is any small number (we use $\epsilon = 10^{-6}$ in the present paper). The resolution Λ_{\min} is the largest Λ_{ij} value for which community detection yields a single community; note that this need not be the minimum non-zero value of Λ_{ij} . Including the small number ϵ in the definition of Λ_{\max} ensures that all edges are antiferromagnetic at resolution $\lambda = \Lambda_{\max}$ and thereby forces each node into its own community.

III. MESOSCOPIC RESPONSE FUNCTIONS (MRFS)

To describe how a network disintegrates into communities as the value of λ is increased from Λ_{\min} to Λ_{\max} (see Fig. 1(a) for a schematic), one needs to select summary statistics. There are many possible ways to summarize such a disintegration process, and we focus on three diagnostics that characterize fundamental properties of network communities.

First, we use the value of the Hamiltonian $\mathcal{H}(\lambda)$ (1), which is a scalar quantity closely related to network modularity and quantifies the energy of the system [13, 14]. Second, we calculate a partition entropy $S(\lambda)$ to characterize the community size distribution. To do this, let n_k denote the number of nodes in community k and define $p_k = n_k/N$ to be the probability to choose uniformly at random a member node of community k . This yields a (Shannon)

partition entropy of $S(\lambda) = -\sum_{k=1}^{\eta(\lambda)} p_k \log p_k$, which quantifies the disorder in the associated community size distribution. Third, we use the number of communities $\eta(\lambda)$.

Needing to normalize \mathcal{H} , S , and η to compare them effectively across networks, we define an *effective energy*

$$\mathcal{H}_{eff}(\lambda) = \frac{\mathcal{H}(\lambda) - \mathcal{H}_{\min}}{\mathcal{H}_{\max} - \mathcal{H}_{\min}} = 1 - \frac{\mathcal{H}(\lambda) - \mathcal{H}_{\min}}{\mathcal{H}_{\max} - \mathcal{H}_{\min}}, \quad (4)$$

where $\mathcal{H}_{\min} = \mathcal{H}(\Lambda_{\min})$ and $\mathcal{H}_{\max} = \mathcal{H}(\Lambda_{\max})$; an *effective entropy*

$$S_{eff}(\lambda) = \frac{S(\lambda) - S_{\min}}{S_{\max} - S_{\min}} = \frac{S(\lambda) - S_{\min}}{\log N}, \quad (5)$$

where $S_{\min} = S(\Lambda_{\min})$ and $S_{\max} = S(\Lambda_{\max})$; and an *effective number of communities*

$$\eta_{eff}(\lambda) = \eta(\lambda) - \eta_{\min} \eta_{\max} - \eta_{\min} = \frac{\eta(\lambda) - 1}{N - 1}, \quad (6)$$

where $\eta_{\min} = \eta(\Lambda_{\min})$ and $\eta_{\max} = \eta(\Lambda_{\max})$.

Some networks contain a small number of entries Λ_{ij} that are orders-of-magnitude larger than most other entries. For example, in the network of Facebook friendships at Caltech [21], 98% of the Λ_{ij} entries are less than 100, but 0.02% of them are larger than 8000. These large Λ_{ij} values arise when two low-strength nodes become connected. Using the null model $P_{ij} = k_i k_j / (2m)$, the interaction between two nodes i and j becomes antiferromagnetic when $\lambda > \Lambda_{ij} / P_{ij} = 2m \Lambda_{ij} / (k_i k_j)$. If the network has a large total edge weight but both i and j have small strengths compared to other nodes in the network, then λ needs to be large to make the interaction antiferromagnetic. In prior studies, network community structure has been investigated at different mesoscopic scale by considering plots of various diagnostics as a function of the resolution parameter [13, 14, 16]. In the present example, such plots would be dominated by interactions that require large resolution-parameter values to become antiferromagnetic. To overcome this issue, we define the *effective fraction of antiferromagnetic edges*

$$\xi = \xi(\lambda) = \frac{\ell^A(\lambda) - \ell^A(\Lambda_{\min})}{\ell^A(\Lambda_{\max}) - \ell^A(\Lambda_{\min})} \in [0, 1], \quad (7)$$

where $\ell^A(\lambda)$ is the total number of antiferromagnetic interactions for the given value of λ in the network. In other words, it is the number of Λ_{ij} elements that are smaller than λ . Thus, $\ell^A(\Lambda_{\min})$ is the largest number of antiferromagnetic interactions for which the network still forms a single community, and the effective number of antiferromagnetic interactions $\xi(\lambda)$ is the number of antiferromagnetic interactions (normalized to the unit interval) in excess of $\ell^A(\Lambda_{\min})$. The function $\xi(\lambda)$ increases monotonically in λ .

Sweeping λ from Λ_{\min} to Λ_{\max} corresponds to sweeping the value of ξ from 0 to 1. (One can think of λ as a continuous variable and ξ as a discrete variable that changes with events.) As we perform such sweeping for a given network, the number of communities increases from $\eta(\xi=0) = 1$ to $\eta(\xi=1) = N$ and yields a vector $(\mathcal{H}_{eff}(\xi), S_{eff}(\xi), \eta_{eff}(\xi))$ whose components we call the *mesoscopic response functions* (MRF) of that network. Because $\mathcal{H}_{eff} \in [0, 1]$, $S_{eff} \in [0, 1]$, $\eta_{eff} \in [0, 1]$, and $\xi \in [0, 1]$ for every network, we can compare the MRFs across networks and use them to identify groups of networks with similar mesoscopic structures. In Fig. 1(b), we show the Zachary Karate Club network [83] for different values of ξ . As more edges become antiferromagnetic, the network fragments into smaller communities, and panel (c) shows the corresponding MRFs. In Fig. 2, we show a schematic of the MRF in which we emphasize its interpretation as a 3-dimensional vector. In Fig. 3, we show example MRFs for several other networks.

Although minimizing Eq. (1) is an NP-hard problem [23] and \mathcal{H} possesses a complicated landscape of local optima for many networks [24], there exist numerous good computational heuristics that make finding a nearly-optimal partition of the network into communities at a given resolution computationally tractable [13, 14]. Thus far, we have reported results that were obtained by optimizing modularity using the locally greedy Louvain algorithm [25] because its speed was important for studying large networks. We have compared the results that we report in the present work to those obtained from optimizing modularity using spectral and simulated-annealing algorithms, and obtained similar MRFs and taxonomies for them (see Appendix B2 for more details).

IV. EXAMPLES OF MRFs

The shapes of the MRFs summarize many factors—including the fraction of possible edges in a network that are actually present, the relative weights of inter- versus intra-community edges, the edge weights compared with the expected edge weights in the null model, the number of edges that need to become antiferromagnetic for a community to fragment, and the way in which the communities fragment (e.g., whether a community splits in half or a single node leaves a community when a particular edge becomes antiferromagnetic). To understand the effects of some of these factors on the shapes of the MRFs, we consider some examples.

Of particular interest are plateaus in the η_{eff} and S_{eff} curves that are accompanied by large increases in \mathcal{H}_{eff} . As illustrated in panel 3(a), the New York Stock HEx -change (NYSE) network from 1980 to 1999 [22] provides a good example of this behavior. This network is an instance from the category of *similarity networks*. We use this label to describe networks that have been constructed by starting from some node-level quantity or attribute and then defining the edges based on some form of similarity or correlation measure between each pair of nodes. Similarity networks tend to be complete (or almost complete) and weighted networks, except when they have been deliberately thresholded. In this particular example, each node represents a stock, and the strength of the edge connecting stocks i and j is linear in the Pearson correlation between the daily logarithmic returns of the stocks. (See Section IX E for more details.) Plateaus imply that as the resolution λ is increased (leading to an increase in \mathcal{H}_{eff}), the communities remain unchanged even though the number and strength of antiferromagnetic interactions increase. As λ is increased and more interactions become antiferromagnetic, there is an increased energy incentive for communities to break up. Community partitions in such plateaus tend to be robust and have the potential to represent interesting structures [13, 14, 16, 27].

In Fig. 3(b), we show MRFs for a “fractal” network [4], which demonstrates that plateaus in the η_{eff} and S_{eff} curves need not be accompanied by significant changes in \mathcal{H}_{eff} . Such plateaus can be explained by considering the distribution of Λ_{ij} values. If several interactions have identical values of Λ_{ij} , then the interactions all become antiferromagnetic at exactly the same resolution value. This leads to a significant increase in the effective fraction of antiferromagnetic edges ξ but only a small change in \mathcal{H}_{eff} . If these interactions do not result in additional communities, then we obtain plateaus in the η_{eff} and S_{eff} curves.

To demonstrate qualitatively different behavior, we show the MRFs for the Biogrid *Drosophila melanogaster* network and the Garfield Scientometrics citation network in Fig. 3(c) and Fig. 3(d), respectively. A common feature in these MRFs is the sharp initial increase in the curves that results from the networks initially breaking into two communities.

Another family of networks, which we will discuss in more detail in our case studies, are political voting networks. These voting networks are also similarity networks: we have constructed these networks so that an edge between two nodes indicates the level of agreement on votes between two entities, and each edge takes a value between 0 and 1. In Fig. 3(e), we show the MRFs for the voting network of the United Kingdom House of Commons during the period 2001–2005 [53]; in Fig. 3(f), we show the MRFs for the roll-call voting network for the 108th (2003–2004) United States House of Representatives [30, 50–52]. In both cases, we observe that sharp increases in \mathcal{H}_{eff} can be accompanied by only small changes in η_e and S_{eff} . To see how this can arise, we again consider the distribution of Λ_{ij} values. If the Λ_{ij} distribution is multi-modal, there can be a large difference between consecutive Λ_{ij} values. A large increase in λ is then needed to increase ξ , which in turn results in a large change in \mathcal{H}_{eff} . However, the change in η_{eff} is small because this only results in a single additional antiferromagnetic interaction.

V. COMPARING NETWORK MODELS

To provide further insights into MRFs, we consider Erdős-Rényi (ER) [1], Barabási-Albert (BA) [3], and Watts-Strogatz (WS) [2] networks. These network models are stochastic, and there is a large ensemble of possible network realizations for each choice of parameter values in these models. However, even with the ensuing structural variation, networks generated by a given one of these three models exhibit similar properties at mesoscopic and macroscopic scales, so we expect MRFs for different realizations of a given model to be similar. In Fig. 4, we compare the MRFs for 1000 realizations of each model for networks with $N = 1000$ nodes and mean degree $\langle k \rangle = 10$. For the WS networks, we set the edge rewiring probability at $p = 0.1$. As illustrated in Fig. 4, we obtain a narrow range of possible MRFs for fixed parameter values. This comparison illustrates that the MRF profiles of the three different models are distinctive. In addition, for each model there is little variation in the behavior of the MRFs across different network realizations with the same parameter values.

It is also instructive to consider variation in MRF shapes for a particular network model for different parameter values. We focus on WS networks because they illuminate the effect of the distribution of Λ_{ij} values on the shapes of the MRFs. In Fig. 5, we show MRFs for WS networks for different values of the edge rewiring probability p . (We continue using $N = 1000$ and $\langle k \rangle = 10$.) We also show the distribution of Λ_{ij} values for each network.

For small rewiring probabilities, the MRFs have lots of steps. As with prior examples, we can see how this feature arises by considering the distribution of Λ_{ij} values. When the rewiring probability is small, many nodes possess the same degree, which results in the presence of many interactions with identical Λ_{ij} values (see the bottom left panel of Fig. 5). Because several interactions have identical Λ_{ij} values, these interactions all become antiferromagnetic at exactly the same resolution-parameter value, so the behavior of MRFs only changes for a small number of ξ values. As the rewiring probability p is increased, the degree and Λ_{ij} distributions become more heterogeneous, which leads to smoother MRFs. For a rewiring probability of $p = 1$, the WS network is just an ER network.

VI. MEASURING DISTANCE BETWEEN NETWORKS

In the framework that we have introduced in this paper, comparing two networks at the mesoscopic level amounts to characterizing the differences in behavior of the corresponding MRFs. To quantify such differences, we define a distance between two networks with respect to one of the summary statistics as the area between the corresponding MRFs. For example, the distance between two networks i and j with respect to the effective energy \mathcal{H}_{eff} is given by

$$d_{ij}^{\mathcal{H}} = \int_0^1 |\mathcal{H}_{eff}^i(\xi) - \mathcal{H}_{eff}^j(\xi)| d\xi. \quad (8)$$

For the effective entropy and effective number of communities, the distances are given by

$$d_{ij}^S = \int_0^1 |S_{eff}^i(\xi) - S_{eff}^j(\xi)| d\xi \text{ and } d_{ij}^\eta = \int_0^1 |\eta_{eff}^i(\xi) - \eta_{eff}^j(\xi)| d\xi, \text{ respectively.}$$

We represent the resulting three sets of distances (computed for each pair of networks from the 746 networks that we consider, see Table I) in matrix form as $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , and \mathbf{D}^η . These distance measures have several desirable properties. First, they compare MRFs across all network scales (i.e., for all values of ξ); second, each distance is bounded between 0 and 1; third, the distances are easy to interpret, as each of them corresponds to the geometric area between (a certain dimension of) a pair of MRFs; and finally, we find *a posteriori* that these distances can be used to cluster networks accurately (see the discussions below).

We have computed MRFs for the energy \mathcal{H} , entropy S , and number of communities η , but we can proceed similarly with any desired summary statistic. If two diagnostics provide similar information, then one of them can be excluded without significant loss of information. We checked whether the summary statistics were sufficiently different, for the set of networks considered here, for it to be worthwhile to include all of them by calculating the Pearson correlation coefficient between their corresponding distance measures. The correlations between the pairs of distances are $r(d_{ij}^{\mathcal{H}}, d_{ij}^S) \doteq 0.36$, $r(d_{ij}^{\mathcal{H}}, d_{ij}^\eta) \doteq 0.24$, and $r(d_{ij}^S, d_{ij}^\eta) \doteq 0.58$. These correlations are not sufficiently high to justify excluding any of the summary statistics.

In the interest of parsimony—and given the non-vanishing correlations between the distance measures—we reduce the number of distance measures using principal component analysis (PCA) [39]. Starting with \mathcal{N} networks, we create a $\frac{1}{2}\mathcal{N}(\mathcal{N}-1) \times 3$ matrix in which each column corresponds to the vector representation of the upper triangle of one of the distance matrices $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , \mathbf{D}^η , and we perform a PCA on this matrix. We then define a distance matrix \mathbf{D}^p with elements $d_{ij}^p = w_{\mathcal{H}} d_{ij}^{\mathcal{H}} + w_S d_{ij}^S + w_\eta d_{ij}^\eta$, where the weights are the coefficients for the first principal component, and we normalize the sum of squared coefficients to unity. The coefficients are $w_{\mathcal{H}} \doteq 0.79$, $w_S = 0.79$, and $w_\eta = 0.57$. The first component accounts for about 69% of the variance, so the distances \mathbf{D}^p provide a reasonable single-variable projection of the distances $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , and \mathbf{D}^η .

It is important that the distance measures for comparing networks are robust to small perturbations in network structure. Because many of the networks that we study are constructed empirically, they might contain false positives and false negatives. In other words, the networks might falsely identify a relationship where none exists, and they also might fail to identify an existing relationship. Consequently, the topology and edge weights of an observed network might be slightly different than those of the actual underlying network. To test the robustness of our distance measures to such observational errors, we recalculate the MRFs for a subset of relatively small unweighted networks in which, for each network, we rewire a number of edges corresponding to a given percentage of the total number of edges (5%, 10%, 20%, 50%, or 100%). See Appendix A for more details. (We study networks with up to 1000 nodes and only consider a subset of 25 networks because of the computational costs of rewiring a large number of networks multiple times; however, we have performed the same investigation for 5 different subsets of 25 networks and obtained similar results. We list the networks in each subset in Table I of the Supplemental Material.) We investigate two rewiring mechanisms: one in which the degree distribution is

maintained, where we also ensure after each rewiring that the network forms a single connected component; and another in which the only constraint is that the network continues to consist of a single connected component after each edge rewiring [40]. We find in both cases that the structures of the block-diagonalized distance matrices for the 25 networks (see Figs. 14 and 15 in Appendix A) are robust to random perturbations of the networks, thereby suggesting that our MRF distance measures are not sensitive to small structural perturbations.

VII. CLUSTERING NETWORKS

We assign each of the 746 networks to a category based on its type (see Table I). Due to the varying availability of different types of network data, the included networks are not evenly distributed across these categories. Many of the networks are either different temporal snapshots of the same system or different realizations of the same type of network. To have a more balanced distribution across the different categories, we focus on 189 of the 746 networks. We only include categories for which we have 8 or more networks, and we selected a subset of networks (uniformly at random) from the larger categories. We also exclude all synthetic networks. See Section IV of the Supplemental Material for the list of networks that we consider and Fig. 1 in Section II of the Supplemental Material for a dendrogram showing a taxonomy we constructed using all 746 networks.

Our primary reason for assigning each network to a category is to use such an external categorization to help assess the quality of taxonomies produced by the unsupervised MRF clustering. For each way of computing distance, we construct a dendrogram for the set of networks using average linkage clustering, which is an agglomerative hierarchical clustering technique [13, 41][42]. In Fig. 6, we show a dendrogram obtained from the distance matrix \mathbf{D}^p . The colored rectangle underneath each leaf indicates the network category. Contiguous blocks of color demonstrate that networks from the same category have been grouped together using the MRF clustering method, and the presence of such contiguous color blocks is an indication of the success of the MRF clustering scheme.

The assignment of the networks to one of these categories is of course to some extent subjective, as several of the networks could belong to more than one category. For example, we could categorize the network of jazz musicians [20] as either a collaboration network or a social network. The initial selection of network categories is also somewhat subjective. One could argue that if one has a social network category, then it is not necessary to have a collaboration network category as well because a collaboration network is a type of social network. We have attempted to maintain a balance between having too many categories and having too few of them. When such ambiguities have arisen, we have systematically chosen the more specific of the relevant categories (e.g., we placed the jazz musician network in the category of collaboration networks rather than in the category of social networks).

VIII. TAXONOMIES OF EMPIRICAL NETWORKS

All of the networks in some categories appear in blocks of adjacent leaves in the dendrogram in Fig. 6. For example, there is a cluster of political voting networks at the far left of the dendrogram. This cluster includes voting networks from the US Senate, the US

House of Representatives, the UK House of Commons, and the United Nations General Assembly (UNGA). The clustering of these voting networks suggests that there are some common features in the network representations of the different legislative bodies. We also obtain blocks that consist of all political committee networks and all metabolic networks.

There are also several categories for which all except one or two networks cluster into a contiguous block. For example, all but two of the fungal networks appear in the same block and all but one of the Facebook networks are clustered together. The isolated Facebook network is the Caltech network, which is the smallest network of this type and which appears in a group next to that containing all of the other Facebook networks. We remark that the social organization of the community structure of the Caltech Facebook network has been shown to be different from those of the other Facebook networks [21].

Networks of certain categories do not appear in near-contiguous blocks. For example, protein interaction networks appear in several clusters. These networks represent interactions within several different organisms, so we would not expect all of them to be clustered together. Moreover, the data that we employed includes examples of protein interaction networks for the same organism in which the interactions were identified using different experimental techniques, and these networks do not cluster together. This supports previous work suggesting that the properties of protein interaction networks are very sensitive to the experimental procedure used to identify the interactions [44, 45]. Social networks are also distributed throughout the dendrogram. This is unsurprising given the extremely broad nature of the category, which includes networks of very different sizes with edges representing a diverse range of social interactions. The leftmost outlying social network is the network of Marvel comic book characters [72], which is arguably an atypical social network.

The grouping (and, to some extent, the non-grouping) of networks by category suggests that the PCA-distance \mathbf{D}^p between MRFs of different networks produces a sensible taxonomy. It is important to ask, however, whether a simpler approach based on a single network diagnostic, such as edge density, can be comparably successful at constructing a taxonomy. In Appendix D, we demonstrate using some well-known diagnostics that this does not appear to be the case, as the diagnostics we tried were unable to reproduce or explain the classifications that we produced using the MRFs.

In order to compare the aggregate shapes of the MRFs across categories, we show the bounds of the \mathcal{H}_{eff} , S_{eff} , and η_{eff} curves for each category in Fig. 7. We again consider all empirical network categories with at least 8 networks in them. This illustrates that the MRFs for some classes of networks (such as political cosponsorship and metabolic networks) are very similar to each other, whereas there are large variations in the MRFs for other categories (such as social and protein interaction networks). The variety of different MRFs for the social and protein interactions is consistent with the fact that their constituent networks are scattered throughout the dendrogram in Fig. 6.

IX. CASE STUDIES

We now consider several case studies, in which we generate taxonomies for multiple realizations of particular types of networks and multiple time slices of particular networks. This enables us to compare these networks and (in some cases) illustrate possible connections between network function and mesoscopic network structure.

A. Voting in the United States Senate

Our first example deals with roll-call voting in the United States Senate [30, 47, 50–52]. Establishing a taxonomy of networks detailing the voting similarities of individual legislators complements previous studies of these data, and it facilitates the comparison of voting similarity networks across time. We consider Congresses 1–110, which cover the period 1789–2008. As in Ref. [50], we construct networks from the roll-call data [30, 51] for each two-year Congress such that the adjacency matrix element $A_{ij} \in [0, 1]$ represents the number of times Senators i and j voted the same way on a bill (either both in favor of it or both against it) divided by the total number of bills on which both of them voted. Following the approach of Ref. [51], we only consider “non-unanimous” roll call votes, which are defined as votes in which at least 3% of the Senators were in the minority.

Much research on the US Congress has been devoted to the ebb and flow of partisan polarization over time and the influence of parties on roll-call voting [50, 52]. In highly polarized legislatures, representatives tend to vote along party lines, so there are strong similarities in the voting patterns of members of the same party and strong differences between members of different parties. In contrast, during periods of low polarization, the party lines become blurred. The notion of partisan polarization can be used to help understand the taxonomy of Senates in Fig. 8, in which we consider two measures of polarization. The first measure uses DW-Nominate scores (a multi-dimensional scaling technique commonly used in political science [51, 52]), where the extent of polarization is given by the absolute value of the difference between the mean first dimension DW-Nominate scores for members of one party and the same mean for members of the other party [30, 51, 52]. In particular, we use the simplest such measure of polarization, called MPR polarization, which assumes a competitive two-party system and hence cannot be calculated prior to the 46th Senate. The second measure we consider is network modularity Q , which was recently shown to be a good measure of polarization even for Congresses without clear party divisions [50]. Modularity is given in terms of the energy \mathcal{H} in Eq. (1) by $Q = -\mathcal{H}(\lambda=1) / (2m)$. These two measures exhibit fairly close agreement on the level of polarization of each Congress for which they can both be calculated [50].

In Fig. 8(a), we include bars under the dendrograms to represent the two polarization measures, both of which have been normalized to lie in the interval $[0, 1]$. The bars demonstrate that Senates with similar levels of polarization (measured in terms of both DW-Nominate scores and modularity values) are usually assigned to the same group, suggesting that our MRF clustering technique groups Senates based on the polarization of roll-call votes. We have also colored dendrogram groups according to their mean levels of polarization using modularity, where the brown group in the dendrogram corresponds to the most highly polarized Senates and the blue group corresponds to the least polarized Senates.

Although one ought to expect similarity in the results from the modularity-based measure of polarization and the MRF clustering, it is important to stress that the MRF clustering method is based on different principles; modularity quantifies the extent to which a given network is “modular”, whereas the MRF clustering explicitly compares the differences in modular structures between any two networks at all scales.

In Fig. 8(a), we also show the clusters that we obtained for the Senate. They closely match the different periods of polarization that have been identified using modularity and DW-Nominate [50]. The cluster with the highest mean polarization (shown in brown) consists of Senates 7, 26–29, 44, 46–51, 53, 55, 66, and 104–110. The 104th–110th Congresses correspond to a period of extremely high polarization following the 1994 “Republican Revolution”, in which the Republican party earned majority status in the House of Representatives for the first time in more than 40 years [30, 50, 52]. The cluster with the second highest mean polarization (shown in red) includes several contiguous blocks of Senates, such as those from Congresses 21–25, 35–39, and 56–61. The 21st–25th Congresses (1829–1839) corresponded to a period of partisan conflict between supporters of John Quincy Adams and Andrew Jackson; it lasted until the emergence of the Whigs and the Democratic party in the 25th Congress [48, 50]. The American Civil War started during the 37th Congress, and a third party known as the Populist Party was strong during the 56th–58th Congresses.

The main differences between different clusters occur in the \mathcal{H}_{eff} response functions. For the most polarized Senates, there is a sharp shoulder in the \mathcal{H}_{eff} MRF that becomes less pronounced as the polarization decreases. We illustrate this in Fig. 9, in which we compare the \mathcal{H}_{eff} MRFs for the (low-polarization) 85th and (high-polarization) 108th Senates. The shoulder in the \mathcal{H}_{eff} curve for the 108th Senate is very pronounced, which can be explained by considering the distribution of Λ_{ij} values. The 108th Senate has a bimodal Λ_{ij} distribution that contains a trough at $\Lambda_{ij} = 1$. Recall that $\Lambda_{ij} = A_{ij}/P_{ij}$, so Λ_{ij} compares the observed voting similarity A_{ij} of legislators i and j with the similarity $P_{ij} = k_i k_j / (2m)$ expected from random voting. If $\Lambda_{ij} < 1$, legislators i and j vote differently more frequently than expected (with respect to the chosen null model); if $\Lambda_{ij} > 1$, they vote more similarly than expected. Therefore, the peaks in the Λ_{ij} distribution above and below 1 correspond, respectively, to intra-party and inter-party voting blocs. In a Senate with low polarization, legislators from different parties often vote in the same manner, so the values of Λ_{ij} no longer separate two distinct types of behavior.

We also examined roll-call voting networks in the US House of Representatives and found many similar features as the ones that we have presented for the US Senate. For example, the highly polarized 104th–110th Congresses, which followed the “Republican Revolution”, appear in the same cluster for both the House and Senate. We also observed some differences in the clusters for the two chambers. For example, the 78th–102nd Senates all appeared in the same cluster. For the House, however, Congresses 80, 88, 89, and 98–102 did not appear in the same cluster as the other Congresses between 78 and 102; instead, they appeared in a cluster that also included the 26th–28th Houses. This was a particularly eventful period: the 25th Congress saw the emergence of the Whigs and the Democratic

Party, and the abolitionist movement was also prevalent (e.g., the Amistad seizure occurred in 1839 during the 26th Congress).

B. Voting in the United Nations General Assembly

The United Nations General Assembly (UNGA) is one of the principal organs of the United Nations (UN), and it is the only part of the UN in which all member nations have equal representation. Although most resolutions are neither legally nor practically enforceable because the General Assembly lacks enforcement powers on most issues, it is the only forum in which a large number of states meet and vote regularly on international issues. It also provides an interesting point of comparison with roll-call voting in the US Congress, as the level of agreement on UN resolutions tends to be much higher than that in the Senate and House [49].

We study voting for the 1st–63rd sessions (covering the period 1946–2008), where each session corresponds to a year [50]. For each session, we define an adjacency matrix \mathbf{A} whose elements A_{ij} represent the number of times countries i and j voted in the same manner in a session (i.e., the sum of the number of times both countries voted yea on the same resolution, both countries voted nay on the same resolution, or both countries abstained from voting on the same resolution) divided by the total number of resolutions on which the UNGA voted in a session. The matrix \mathbf{A} , with elements $A_{ij} \in [0, 1]$, thereby represents a (similarity) network of weighted edges between countries.

We cluster UNGA sessions by comparing MRFs for the corresponding voting networks. In Fig. 10, we plot a dendrogram of the UNGA sessions and highlight some of the clusters, which correspond to notable periods in the recent history of international relations. The red cluster in the middle of the dendrogram consists of all post-Cold War sessions (1992–2008) except 1995. This group forms a larger cluster with some UNGA sessions from the 1970s and a cluster consisting of 1946, 1948, and 1950. These three sessions (shown in magenta) are all noteworthy: 1946 was the first session of the UNGA, the Universal Declaration of Human Rights was introduced during the 1948 session, and the “Uniting for Peace” resolution was passed during the 1950 session. At the rightmost part of the dendrogram, we color in black a group that consists of all sessions from 1979 to 1991 (excluding 1980). The beginning of this period marked the end of Détente between the Soviet Union and the US following the former's invasion of Afghanistan at the end of 1979, and the end of this period saw the end of the Cold War. The large blue cluster in the leftmost part of the dendrogram consists primarily of sessions from before 1971 (though it also includes the sessions in 1977 and 1995).

C. Facebook

We now consider Facebook networks for 100 US universities [21]. The nodes in each network represent users of the Facebook social networking site, and the unweighted edges represent reciprocated “friendships” between users at a single-time snapshot in September 2005. We consider only edges between students at the same university, as this allows us to compare the structure of the networks at the different institutions. These networks represent complete data sets obtained directly from Facebook. In contrast to the previous examples,

we are not comparing snapshots of the same network at different times but are instead comparing multiple realizations of the same type of network that have evolved independently. Such real-world ensembles of network data are rare, and constructing a taxonomy will hopefully allow us to compare and contrast the social organization at these institutions.

In Fig. 11, we show the dendrogram for Facebook networks that we produced by comparing MRFs. The two color bars below the dendrogram indicate (top) the number of nodes N in each network and (bottom) the fraction of possible edges d that are present (i.e., edge density). The Facebook networks range in size from 762 to 41,536 nodes, and the edge density varies from 0.2% to 6%. In contrast to previous examples, we observe in this case that two simple network properties appear to explain most of the observed clustering of the networks. An important feature of this example is that the \mathcal{H}_{eff} , S_{eff} , and η_{eff} MRFs are each very similar in shape and lie in a narrow range across all 100 institutions (see Fig. 7). Such extreme similarity is remarkable—as one can see in Fig. 7, this contrasts starkly with most of the other examples—and it suggests that all of the Facebook networks have very similar mesoscopic structural features. If one also considers demographic information, then one can find interesting differences between the networks [21], but the structural similarity is striking.

D. Fungi

We also examined fungal mycelial networks extracted from time series of digitized images of colony growth. In these undirected, planar, weighted networks, the nodes represent hyphal tips, branch points, or anastomoses (hyphal fusions), and the edges represent the interconnecting hyphal cords weighted by their conductivity [27, 52, 53]. For comparison, we also digitized weighted networks of the acellular slime mold *Physarum polycephalum* [24]. Fungal networks look like trees but contain additional edges (known as *cross-links*) that generate cycles.

As shown in Fig. 12(a), we find using our method that replicate networks from different species at comparable time points are grouped together. Furthermore, the aggregate clustering pattern reflects increasing levels of cross-linking that are characteristic of different species, as illustrated in Fig. 12(b); this ranges from the low levels in *Resinicium bicolor* to intermediate levels in *Phanerochaete velutina* and highly cross-linked networks formed by *Phallus impudicus*. By constructing a dendrogram for only one species but including data from repeated experiments and over time (see Fig. 12(c)), we observe a progression from trees at early developmental times to an increasingly cross-linked network later in mycelium growth [26, 27]. In early growth, the developmental stage appears to dominate the clustering pattern, as networks from different replicates but of similar age are grouped together. At later times, however, networks show a high aggregate level of similarity, and the fine-grained clustering predominantly reflects the subtle changes in structure evolving within each replicate.

E. New York Stock Exchange

As our final example, we consider a set of stock-return correlation networks for the New York Stock Exchange (NYSE), which is the largest stock exchange in the world (as measured by the aggregate US dollar value of the securities listed on it). Each node represents a stock, and the strength of the edge connecting stocks i and j is linear in the Pearson product-moment correlation coefficient between the daily logarithmic returns of the stocks [22]. We consider $N = 100$ stocks during the time period 1985–2008 and construct a network for each 6 months of data. This yields a sequence of fully-connected, weighted adjacency matrices whose elements quantify the similarity of two stocks (normalized to the unit interval for each time window).

We show the dendrogram for the NYSE networks in Fig. 13. The first division of these networks classifies them into two groups (which we have colored in blue and red). The red cluster appears to correspond to periods of market turmoil, including the networks for the second half of 1987 (including the Black Monday crash of October 1987), all of 2000–2002 (including and following the bursting of the dot-com bubble), and the second half of 2007 and all of 2008 (including the recent credit and liquidity crisis). The value of the NYSE composite index, which measures the aggregate performance of all common stocks listed on the NYSE [56], supports our hypothesis that the red cluster is associated with periods of market turmoil. Indeed, the networks in the red cluster correspond (with one or two exceptions) to the periods of high volatility of the composite index (see Fig. 13).

X. CONCLUSIONS

We have developed an approach that facilitates the comparison of diverse networks by summarizing network community structure using what we call mesoscopic response functions (MRFs). We have demonstrated how this approach can be used to group networks both across categories and within categories. Our work builds on prior research on network community structure, which has focused predominantly on algorithmic detection of the communities rather than on subsequently using the communities for applications (such as comparing sets of networks).

The development of algorithmic methods to detect communities is frequently motivated by the idea that the community structure of a network representing a system has some bearing on the function of the system. If different networks perform different functions—and if their functions are constrained, at least in part, by their mesoscopic structure—then it should be possible in principle to derive a functional classification of networks based on community structure. Although this has mostly been presented as a presumption in the existing literature, it is actually an empirically testable hypothesis. Indeed, we have shown in the present paper that one can systematically exploit mesoscopic structure to obtain useful comparisons of networks. This allows one to derive taxonomies for networks that also appear to have correspondence with functional similarities. We observed that networks that were not grouped with other members of the same class appeared to be unusual in some respects, and we also demonstrated that we could detect historically-noted financial and political changes from time-ordered sequences of networks.

We believe that our framework has the potential to aid in the exploration and exploitation of similarities in network structures across both network types and disciplinary boundaries.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Each of the networks that we have used is described in detail in the Supplemental Material, where we have also provide relevant citations for the data sources. We are particularly grateful to A. D'Angelo and Facebook, A. Merdzanovic, M. E. J. Newman, E. Voeten, and Vote-view [30] (maintained by K. Poole and H. Rosenthal) for data. We thank A. Lewis, G. Villar, S. Agarwal, D. Smith, and the anonymous referees for useful suggestions. We adapted code originally written by R. Guimerà for the simulated annealing algorithm and code by R. Lambiotte for the Louvain algorithm. This research was supported in part by the Fulbright Program (JPO) and NIH grant P01 AG-031093 (JPO), the James S. McDonnell Foundation (MAP: #220020177), NSF (PJM: DMS-0645369), and EPSRC and BBSRC (NSJ: EP/I005765/1, EP/I005986/1, and BBD0201901).

Appendix A: Robustness of Clustering

To examine the robustness of our clustering to false positives (false links) and false negatives (false non-links), we consider two network rewiring mechanisms, and we apply the rewiring to each network in a subset of 25 networks highlighted in Table 2 of the Supplemental Material. The first step in the procedure is to randomly rewire a number of edges corresponding to a given percentage (5%, 10%, 20%, 50%, or 100%) of the total number of edges in the network, subject to the constraints that we preserve the network's degree distribution and the fact that it consists of a single connected component [57]. (That is, such a rewiring of a number of edges equal to $x\%$ of the L edges in a network means that we perform $\lceil xL \rceil$ rewiring steps; the same edge can be rewired multiple times.) Second, we randomly rewire a given number of the edges subject only to the constraint that the rewired network still consists of a single component.

Because we are perturbing the original network, we focus on the distance matrices $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , and \mathbf{D}^{η} as they can be calculated directly for each network. We consider 25 of the 746 original networks of varying sizes and edge densities; we highlight these networks in bold in Table II of the Supplemental Material. In Fig. 14, we show the distance matrices for this subset of networks when different percentages of edges have been rewired with the degree distribution preserved. The first column shows the matrices for the original networks. (Note that the node orderings for $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , and \mathbf{D}^{η} are not necessarily the same in Fig. 14 because of the block-diagonalization of matrices.) The subsequent columns show the mean distance matrices as increasing numbers of edges are rewired; for a given row, the node ordering in each column is fixed. The distance matrices for the randomizations are the mean pairwise distances between networks, where the mean is calculated over all possible pairs between 10 perturbations of each network. More precisely, let A and B represent two different (unperturbed) networks and let the sequences A_1, A_2, \dots, A_{10} and B_1, B_2, \dots, B_{10} represent 10 realizations of the perturbation process (e.g., at the 5% level) for the networks. To calculate the distance between A and B under perturbation, we find for each $j \in \{1, \dots, 10\}$ the distances between A_j and B_1 , A_j and B_2, \dots , and A_j and B_{10} . We then calculate the mean of the ensuing $10 \times 10 = 100$ distance values. Based on visual inspection of Fig. 14, the

matrices for the first few columns for all of the distances are fairly similar to the original distance matrices. This suggests some notion of robustness in our clustering technique. We study only 25 networks because of the computational costs of rewiring a large number of networks multiple times; however, we have performed the same investigation for 5 different subsets of 25 networks and obtained similar results. We list the networks in each subset of 25 in Table I in the Supplemental Material.

To carry out a more thorough randomization of each network, we now rewire every edge in the network 10 times on average. In Fig. 15, we show the $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , and \mathbf{D}^η mean-distance matrices for this number of rewirings. We again calculate the mean distance using the method described in the previous paragraph. The first column again shows the distance matrices for the original networks. The second and third columns show the distance matrices for randomizations in which the degree distribution is preserved and destroyed, respectively. The node orderings of the matrices in the second and third columns are again the same as the orderings for the matrix of the first column of the corresponding row. The second column in Fig. 15 demonstrates that some block structure remains in the distance matrices when the degree distribution is preserved. The third column shows that much of this structure is destroyed (though some block structure is still visible) when the degree distribution is not preserved. When the networks are “fully randomized” in this way—with the only constraint being that each rewired network must consist of a single connected component—one is in effect producing random graphs. These random graphs might, however, still have some common properties, such as the number of nodes and the edge density.

Appendix B: Computational Heuristics

1. Robustness of Network MRFs

We detected all communities in the main text using the locally greedy Louvain algorithm [25]; however, several alternative heuristics exist, so we now investigate whether the choice of heuristic has any effect on the results. In Ref. [24], Good et al. demonstrated that there can be extreme near-degeneracies in the energy function, in particular an exponential number of low-energy (i.e., high-modularity) solutions. Given this, it is unsurprising that different energy-optimization heuristics can yield very different partitions for the same network. Good et al. suggested that the reason for this behavior is that different heuristics sample different regions of the energy landscape. Because of the potential sensitivity of results to the choice of heuristic, one should treat individual partitions by particular heuristics with caution. However, one can have more confidence in the validity of the partitions if different heuristics produce similar results. Here we compare the results for the Louvain algorithm [25] with those for a spectral algorithm [18] and simulated annealing [58].

In Fig. 16, we show MRFs for three networks calculated using Louvain [25], spectral [18] and simulated annealing algorithms [58]. For all three networks, the three algorithms agree very closely on the shapes of the \mathcal{H} , S , and η MRFs. The MRFs are most similar for the roll-call voting network of the 102nd US Senate [50–52], and the \mathcal{H} MRF is almost identical for the three heuristics. In general, we observe the largest differences in the shapes

of the MRFs when using the spectral algorithm. The spectral algorithm that we used begins by finding a partition of the network into exactly two components such that the energy is minimized (among all bipartitions). It then recursively partitions the smaller networks into two groups until no decrease in energy can be obtained through bipartitioning. At each step, this algorithm only finds the optimal partition of each community into two smaller communities even though a split into more communities could yield a lower energy. Given this, it is unsurprising that the spectral algorithm often identifies partitions further from the optimum than the other heuristics. For the remainder of this section, we therefore only compare the Louvain and simulated annealing algorithms.

2. Robustness of Resulting Network Taxonomies

Although Fig. 16 shows good agreement between the shapes of the MRFs that we obtain from the different computational heuristics, we nevertheless check that the small differences that do occur do not have a significant effect on the resulting network taxonomy. Because of the computational cost of detecting communities using simulated annealing, we investigate the effect on the taxonomy using a subset of small networks. We highlight all of the networks that we consider with an asterisk (*) in Table II of the Supplemental Material. (The largest network that we include is the cat brain cortical/thalamic network [11], which has 1,170 nodes.). Indeed, MRFs for small networks tend to be much noisier than those for large networks—see, for example, Fig. 16(a), which shows the MRFs for the 34-node Zachary Karate Club network—so we expect that any differences between algorithms are likely to be more pronounced for small networks.

In Fig. 17, we show dendrograms obtained using the Louvain and simulated-annealing modularity optimization algorithms for a subset of 15 networks. On visual inspection, the dendrograms appear to be very similar, as there are only a few small differences in the heights at which leaves and clusters combine. To quantify the similarity between a pair of dendrograms with underlying distance matrices \mathbf{s} and \mathbf{t} , we define a correlation coefficient ϕ as

$$\phi = \frac{\sum_{i>j} (s_{ij} - \bar{s}) (t_{ij} - \bar{t})}{\sqrt{\left[\sum_{i>j} (s_{ij} - \bar{s})^2 \right] \left[\sum_{i>j} (t_{ij} - \bar{t})^2 \right]}}, \quad (\text{B1})$$

where \bar{s} is the mean of the distances s_{ij} and \bar{t} is the mean of the distances t_{ij} . Dendrograms derived from identical distance matrices have correlation coefficient $\phi = 1$. The correlation for the example dendrograms shown in Fig. 17 is 0.997. To get a better sense of the extent of this correlation, we compare the observed correlations with those obtained for randomized dendrograms. To make the comparison, we first produce a distribution of correlation coefficients ϕ between a large number of empirical (unrandomized) dendrograms produced by the Louvain and simulated-annealing algorithms. Because of the computational costs of calculating the MRFs for the simulated annealing algorithm, we only consider the subset of 25 networks identified above. We select 15 networks uniformly at random from this subset of 25 networks and generate two dendrograms similar to those in Fig. 17: one corresponds to

the distance matrix produced by the Louvain algorithm and the other corresponds to the distance matrix produced by simulated annealing. We then calculate the correlation coefficient between the two distance matrices. We repeat this process 10,000 times to obtain 10,000 correlation coefficients, whose distribution we show using the hollow red histogram in Fig. 18. This procedure makes it possible to compare a large number of dendrograms at the computational cost of calculating simulated annealing MRFs for a total of 25 networks, highlighted with asterisks in Table 2 of the Supplemental Material.

We then compare this observed distribution of correlation coefficients to a randomized reference. We focus on the correlation between empirical Louvain dendrograms (i.e., empirical dendrograms resulting from distance matrices produced by the Louvain method) and randomized simulated-annealing dendrograms (i.e., dendrograms resulting from distance matrices produced by the simulated annealing algorithm that have been subsequently randomized). We proceed as follows: for each of the 10,000 dendrogram pairs that we assembled from subsets of 15 networks, we create 100 randomizations of the simulated-annealing dendrogram, and we then calculate the correlation coefficient between each of these randomized dendrograms and the corresponding empirical Louvain dendrogram. The resulting distribution from 10,000 repetitions is the solid blue histogram in Fig. 18. To randomize the simulated-annealing dendrogram, we used the double-permutation procedure described in Refs. [60, 61]. This procedure has two steps. First, we randomize the distances at which the different clusters are combined. For example, consider an unrandomized dendrogram in which clusters A and B are combined at a distance of 0.45 and clusters C and D are combined at a distance of 0.65; after the randomization, A and B might be combined at a distance of 0.65 and C and D might be combined at a distance of 0.45. Second, we randomize the networks corresponding to each leaf in the dendrogram. This two-step randomization procedure maintains the underlying distances and the topology of the dendrogram.

As mentioned above, we show the distributions of correlation coefficients between empirical Louvain dendrograms and the empirical (unrandomized) and randomized simulated-annealing dendrograms in Fig. 18. The correlation is clearly much higher for the empirical case, as there is only a very slight overlap in the tails of the two distributions. The correlation between the Louvain and simulated-annealing dendrograms is greater than 0.99 for about 63% of the studied dendrograms.

Appendix C: Diagnostic for Assessing the Clustering from Different Distance Measures

An examination of the leaf colors of the dendrogram in Fig. 7 illustrates that the employed distance measure groups together networks from a variety of categories, including political voting networks, political committee networks, Facebook networks, metabolic networks, and fungal networks. A visual comparison provides a reasonable starting point for assessing the effectiveness of different distance measures at clustering networks. To quantify how effectively each distance matrix (\mathbf{a} , \mathbf{D}^S , \mathbf{D}^η , and \mathbf{D}^p) clusters networks of the same type, we introduce a *clustering diagnostic*, which we denote by $\alpha(h)$, to be explained shortly. Because

the assignment of networks to categories is subjective and because some of the categories include networks of very different types, it would be inappropriate to assess the effectiveness of a distance measure based on how well it clusters networks in very broad categories. We thus focus our examination on narrower categories whose constituent networks are clustered fairly tightly in Fig. 7. This includes the following 8 categories of networks: Facebook, metabolic, political cosponsorship, political committee, political voting, financial, brain, and fungal.

The clustering diagnostic depends on where one “cuts” the dendrograms. We start by constructing a dendrogram for each of the four distance matrices $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , \mathbf{D}^η , and \mathbf{D}^p . Performing a horizontal cut through a dendrogram at a given height h splits the dendrogram into multiple disconnected clusters (h is measured in terms of ultrametric distances; see Fig. 17). For each such cluster, we calculate the proportion of networks from a particular category that are contained in it. For example, if a cut produces three clusters and if we consider the Facebook category, then we might find that one cluster contains two tenths of the Facebook networks, a second cluster has three tenths of those networks, and the third cluster has the remaining half of those networks. We calculate these membership fractions for each network category and for each cluster. We then identify, for each category, what we called the *plurality cluster*, which is defined as the cluster that includes the largest fraction of networks from that category. In the above example, the third cluster is the plurality cluster for the Facebook category. Our diagnostic $\alpha(h)$ is then defined by adding across all 8 categories the fraction of networks in the plurality clusters:

$$\alpha(h) = \sum_{j=1}^8 \gamma_j(h), \quad (\text{C1})$$

where $\gamma_j(h)$ is the plurality fraction for the j th category of networks for the given cut at height h of the dendrogram.

We perform similar calculations for each level of the dendrogram and use the resulting values of $\alpha(h)$ to assess the effectiveness of the different distance measures at clustering the networks. For example, at the root of the dendrogram, all of the networks are in a single cluster, so the maximum fraction of networks in the same cluster is 1 for every network category. Given the above choice of 8 categories, this yields $\alpha = 8$. However, as one considers lower levels of the dendrogram, the clusters break up more and more, so the fraction of networks in the plurality cluster in each category typically decreases. Effective distances measures ought to result in relatively high values for $\alpha(h)$.

In Fig. 19, we compare the values of $\alpha(h)$ at each level of the dendrogram for $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , \mathbf{D}^η , and \mathbf{D}^p . For each of the different subsets of networks and for most of the dendrogram levels, the PCA-distance \mathbf{D}^p is the most effective of the employed distance measures at clustering networks of the same category. This agrees with our visual assessment (i.e., our identification of contiguous blocks of color) of the different measures.

Appendix D: Using Simple Characteristics to Cluster Networks

We established in Section VIII that the PCA-distances \mathbf{D}^p between MRFs can produce sensible network taxonomies, and we now consider briefly whether the observed taxonomies can be explained using simple summary statistics. We consider only a few specific properties, though of course there are myriad other network diagnostics that one might consider.

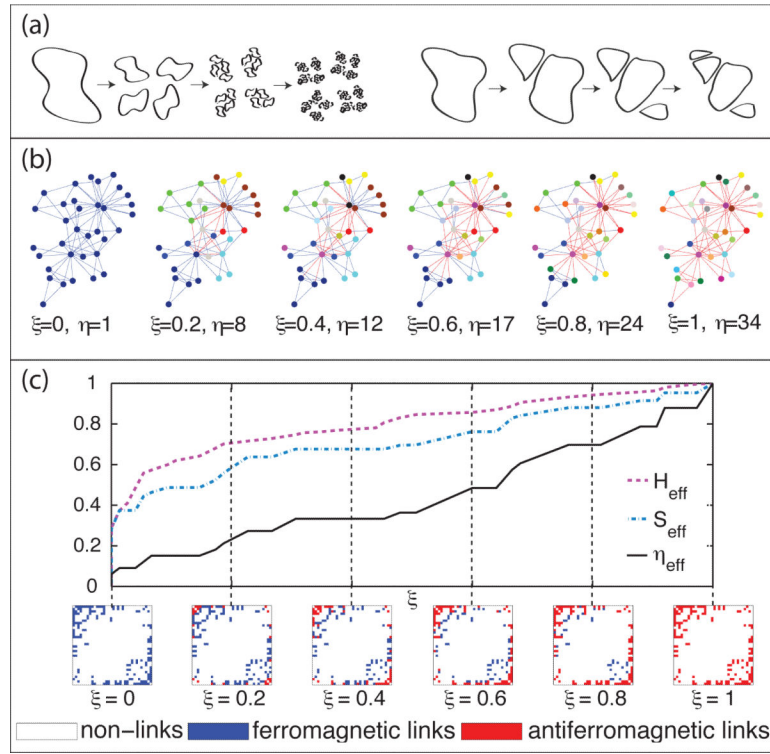
Perhaps the three simplest properties of an undirected network are the following: (1) whether it has weighted or unweighted edges; (2) the number of nodes N ; and (3) the edge density $d = 2L/[N(N-1)]$ (where L is the number of edges, which we distinguish from the total edge weight m in weighted networks). The top colored row in Fig. 20 indicates that many of the weighted networks are clustered together at the far left of the dendrogram. However, there are also weighted networks scattered throughout the dendrogram, so whether a network is weighted or unweighted does not explain the observed classification. The third colored row provides a clearer explanation for the cluster of networks at the left: These are not simply weighted networks, as they are in fact similarity networks, so that nearly all possible edges are present and have weights indicating connection strengths. However, this property alone cannot explain the observed classification, as several of the weighted networks containing nearly all possible edges do not appear at the far left of the dendrogram. In fact, there are many clusters in the dendrogram that contain networks with very different fractions of possible edges. The total number of nodes, shown by the second colored row in the figure, again explains some of the clustering, as networks with similar numbers of nodes are clustered together in some regions of the dendrogram. However, there are also numerous examples in which networks with the same number of nodes appear in different clusters. Therefore, none of these three simple network diagnostics can explain the observed classification by itself.

References

1. Newman, MEJ. *Networks: An Introduction*. Oxford University Press; Oxford, UK: 2010.
2. Newman MEJ. *Physics Today*. 2008; 61:33.
3. Börner, K. *Atlas of Science: Visualizing What We Know*. MIT Press; Cambridge, MA, USA: 2010.
4. The code used for constructing network taxonomies in this paper is available at www.jponnola.com.
5. Albert R, Barabási A-L. *Review of Modern Physics*. 2002; 74:47.
6. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. *Science*. 2002; 298:824. [PubMed: 12399590]
7. Milo R, Itzkovitz S, Kashtan N, Levittand R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U. *Science*. 2004; 303:1538. [PubMed: 15001784]
8. Guimerà R, Sales-Pardo M, Amaral LAN. *Nature Physics*. 2007; 3:63.
9. Serrano MA, Boguná M, Vespignani A. *Proceedings of the National Academy of Sciences of the U.S.A.* 2009; 106:6483.
10. Arenas A, Borge-Holthoefer J, Gomez A, Zamora G. *New Journal of Physics*. 2010; 12:053009.
11. Bagrow J, Bollt E, Skufca J, Ben-Avraham D. *EPL (Europhysics Letters)*. 2008; 81:68004.
12. Song C, Havlin S, Makse HA. *Nature*. 2005; 433:392. [PubMed: 15674285]
13. Porter MA, Onnela J-P, Mucha PJ. *Notices of the American Mathematical Society*. 2009; 56:1082.
14. Fortunato S. *Physics Reports*. 2010; 486:75.

15. Lancichinetti A, Kivela M, Saramaki J, Fortunato S. PLoS One. 2010; 5:e11976. [PubMed: 20711338]
16. Reichardt J, Bornholdt S. Physical Review E. 2006; 74:016110.
17. Newman MEJ, Girvan M. Physical Review E. 2004; 69:026113.
18. Newman MEJ. Proceedings of the National Academy of Sciences of the U.S.A. 2006; 103:8577.
19. Reichardt J, Bornholdt S. Physical Review Letters. 2004; 93:218701. [PubMed: 15601068]
20. Traud AL, Kelsic ED, Mucha PJ, Porter MA. SIAM Review. 2011; 53:526.
21. Traud AL, Kelsic ED, Mucha PJ, Porter MA. arXiv:1102.2166. 2011
22. Zachary WW. Journal of Anthropological Research. 1977; 33:452.
23. Brandes U, Dellling D, Gaertler M, Görke R, Hoefer M, Nikoloski Z, Wagner D. IEEE Transactions on Knowledge and Data Engineering. 2008; 20:172.
24. Good BH, de Montjoye YA, Clauset A. Physical Review E. 2010; 81:46106.
25. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Journal of Statistical Mechanics. 2008; 2008:P10008.
26. Onnela J-P, Chakraborti A, Kaski K, Kertész J, Kanto A. Physical Review E. 2003; 68:056110.
27. Arenas A, Fernandez A, Gómez S. New Journal of Physics. 2008; 10:053039.
28. Sporns O. BioSystems. 2006; 85:55. [PubMed: 16757100]
29. Firth, D.; Spirling, A. Working paper. 2003. available at <http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic/firth/software/tapir/firth-spirling.pdf>
30. Poole, KT. 2009. Data available at <http://voteview.com>
31. Poole, KT.; Rosenthal, H. Congress: A Political-Economic History of Roll Call Voting. Oxford University Press; Oxford, UK: 1997.
32. McCarty, NM.; Poole, KT.; Rosenthal, H. Polarized America: The Dance of Ideology and Unequal Riches. MIT Press; Cambridge, MA, USA: 2007.
33. Waugh AS, Pei L, Fowler JH, Mucha PJ, Porter MA. arXiv:0907.3509. 2010
34. Stark C, Breitkreutz BJ, Regul T, Boucher L, Breitkreutz A, Tyers M. Nucleic Acids Research. 2006; 34:D535. [PubMed: 16381927]
35. Garfield, E.; Sher, IH.; Torpie, RJ. The Use of Citation Data in Writing the History of Science. The Institute for Scientific Information; Philadelphia, PA, USA: 1964.
36. Erdős P, Rényi A. Publicationes Mathematicae. 1959; 6:290.
37. Barabási A-L, Albert R. Science. 1999; 286:509. [PubMed: 10521342]
38. Watts DJ, Strogatz SH. Nature. 1998; 393:440. [PubMed: 9623998]
39. Jolliffe, IT. Principal Component Analysis. Springer-Verlag; New York, NY, USA: 1986.
40. Maslov S, Sneppen K. Science. 2002; 296:910. [PubMed: 11988575]
41. Duda, RO.; Hart, PE.; Stork, DG. Pattern Classification. Wiley; New York, NY, USA: 2001.
42. The three most common linkage clustering algorithms are single, average, and complete linkage clustering; they join clusters of objects based on the smallest, mean, and largest distance between objects in the clusters, respectively. We used the cophenetic correlation ζ to compare how well each algorithm preserves the pairwise distances between the networks [?]. We obtain the following cophenetic correlations . for the different linkage clustering algorithms: $\zeta_{\text{single}} = 0.65$, $\zeta_{\text{avg}} = 0.78$, and $\zeta_{\text{comp}} = 0.62$. Hence, dendrograms constructed using average linkage clustering preserve the distances in D_p better than those constructed using the other two clustering techniques. We thus use average linkage clustering to construct all of the dendrograms in this paper.
43. Gleiser P, Danon L. Advances in Complex Systems. 2003; 6:565.
44. Hakes L, Pinney JW, Robertson DL, Lovell SC. Nature Biotechnology. 2008; 26:69.
45. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, et al. Science. 2008; 322:104. [PubMed: 18719252]
46. Alberich R, Miro-Julia J, Rossello F. arXiv:condmat/0202174. 2002
47. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela J-P. Science. 2010; 328:876. [PubMed: 20466926]

48. Kernell, S.; Jacobson, GC.; Kousser, T. *The Logic of American Politics*. 4th ed.. CQ Press; Washington, DC, USA: 2009.
49. Macon KT, Mucha PJ, Porter MA. *Physica A*. 2012; 391:343.
50. We excluded the 19th session because only one resolution was considered by the UNGA that year.
51. Bebbler DP, Hynes J, Darrah PR, Boddy L, Fricker MD. *Proceedings of the Royal Society B*. 2007; 274:2307. [PubMed: 17623638]
52. Fricker, M.; Boddy, L.; Bebbler, D. Network organization of mycelial fungi. In: Howard, RJ.; Gow, NAR., editors. *The Mycota. Vol VIII, Biology of the Fungal Cell*. Springer; Berlin, Germany: 2007. p. 307-328.
53. Fricker, MD.; Boddy, L.; Nakagaki, T.; Bebbler, DP. Adaptive biological networks. In: Gross, T.; Sayama, H., editors. *Adaptive Networks: Theory, Models and Applications*. Springer; Berlin/Heidelberg, Germany: 2009. p. 51-70.
54. Tero A, Takagi S, Saigusa T, Ito K, Bebbler DP, Fricker MD, Yumiki K, Kobayashi R, Nakagaki T. *Science*. 2010; 327:439. [PubMed: 20093467]
55. Heaton LLM, López E, Maini PK, Fricker MD, Jones NS. *Proceedings of the Royal Society B*. 2010; 277:3265. [PubMed: 20538649]
56. See <http://www.nyse.com/>.
57. Maslov S, Sneppen K. *Science*. 2002; 296:910. [PubMed: 11988575]
58. Guimerà R, Sales-Pardo M, Amaral LAN. *Physical Review E*. 2004; 70:025101.
59. Scannell JW, Burns GAPC, Hilgetag CC, O'Neil MA, Young MP. *Cerebral Cortex*. 1999; 9:277. [PubMed: 10355908]
60. Lapointe F-J, Legendre P. *Journal of Classification*. 1995; 12:265.
61. Lapointe F-J, Legendre P. *Systematic Zoology*. 1990; 39:1.

**FIG. 1.**

(Color online) (a) Schematic of some of the ways that a network can break up into communities as the value of λ (or ξ) is increased. (b) Zachary Karate Club network [83] for different values of the effective fraction of antiferromagnetic edges ξ . All interactions are either ferromagnetic or antiferromagnetic, i.e. for the values of ξ used, there are no neutral interactions. We color edges in blue if the corresponding interactions are ferromagnetic, and we color them red if the interactions are antiferromagnetic. We color the nodes based on community affiliation. (c) The \mathcal{H}_{eff} , S_{eff} , and η_{eff} MRFs, and the interaction matrix \mathbf{J} for different values of ξ . We color elements of the interaction matrix by depicting the absence of an edge in white, ferromagnetic edges in blue, and antiferromagnetic edges in red.

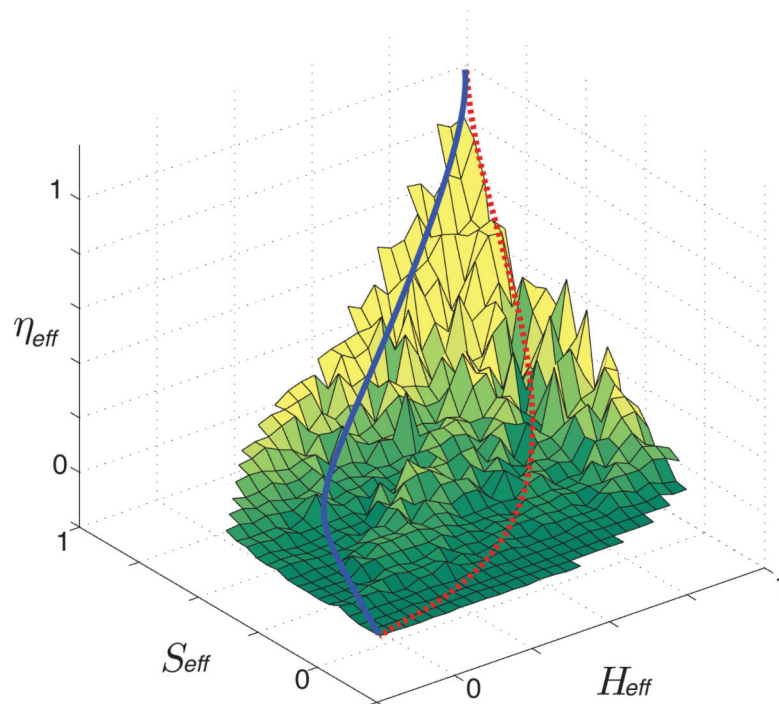


FIG. 2.

(Color online) The mesoscopic response function (MRF) of a given network consists of a 3-dimensional vector $(\mathcal{H}_{eff}(\xi), S_{eff}(\xi), \eta_{eff}(\xi))$, where $\xi \in [0, 1]$. By construction, the MRF starts from the bottom front corner $[\mathcal{H}_{eff}(\xi=0), S_{eff}(\xi=0), \eta_{eff}(\xi=0)]$ and ends at the top back corner $[\mathcal{H}_{eff}(\xi=1), S_{eff}(\xi=1), \eta_{eff}(\xi=1)]$. The colored surface plot shows where most MRFs lie. We also show schematic MRFs in blue (solid curve) and red (dashed curve).

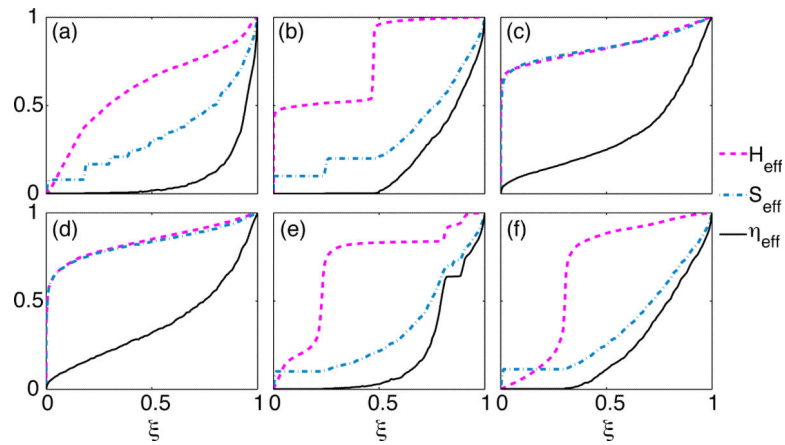


FIG. 3.

(Color online) Example mesoscopic response functions (MRFs). The curves show \mathcal{H}_{eff} (pink, dashed), S_{eff} (blue, dash-dotted), and η_{eff} (black, solid) as a function of the effective fraction of antiferromagnetic edges ξ for the following networks: **(a)** New York Stock Exchange (NYSE), 1980–1999 [22]; **(b)** Fractal (10,2,8) [4]; **(c)** Biogrid *D. melanogaster* [55]; **(d)** Garfield scientometrics citations [40]; **(e)** United Kingdom House of Commons voting, 2001–2005 [53]; **(f)** Roll-call voting of 108th United States House of Representatives [30, 50–52].

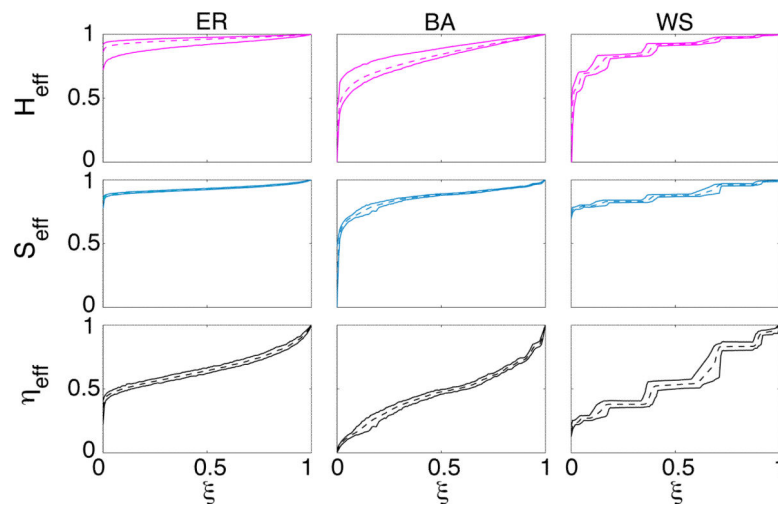


FIG. 4. (Color online) MRFs for 1000 realizations of Erdős-Rényi (ER), Barabási-Albert (BA), and Watts-Strogatz (WS) networks. Each network has $N = 1000$ nodes and mean degree $\langle k \rangle = 10$. For each value of ξ , the upper curves show the maximum values of \mathcal{H}_{eff} (top row), S_{eff} (middle row), and η_{eff} (bottom row) for all networks in the ensemble; the lower curves show the corresponding minimum value, and the dashed curves show the corresponding mean.

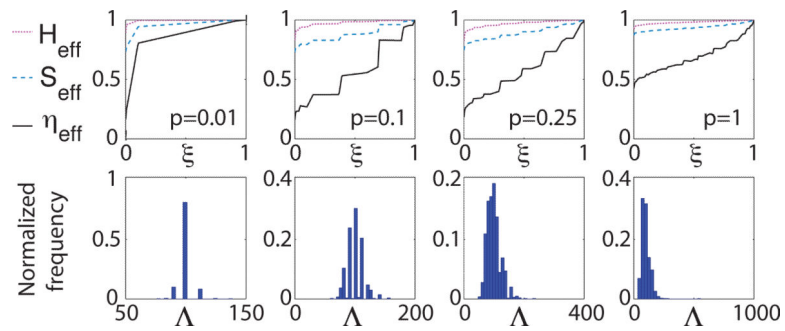
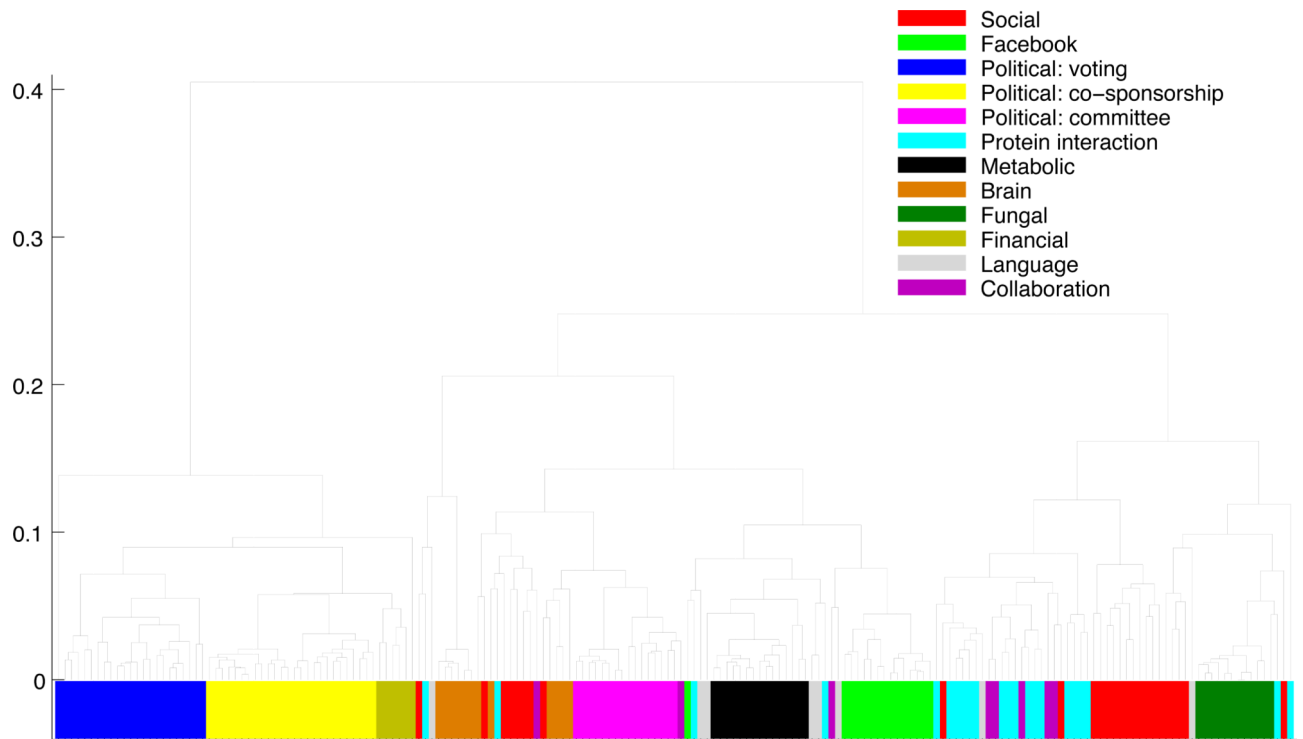


FIG. 5.

(Color online) Upper panels: MRFs for Watts-Strogatz networks for different values of the rewiring probability p . Each network has $N = 1000$ nodes and mean degree $\langle k \rangle = 10$. Lower panels: distributions of Λ_{ij} values for each network. As expected, the MRFs for $p = 1$ are identical to those of an Erdős-Rényi network with $N = 1000$ and $\langle k \rangle = 10$.

**FIG. 6.**

(Color online) Taxonomy for 189 networks. We construct the dendrogram (tree) using the distance \mathbf{D}^p and average linkage clustering. We order the leaves of the dendrogram to minimize the distance between adjacent nodes and color the leaves to indicate the type of network.

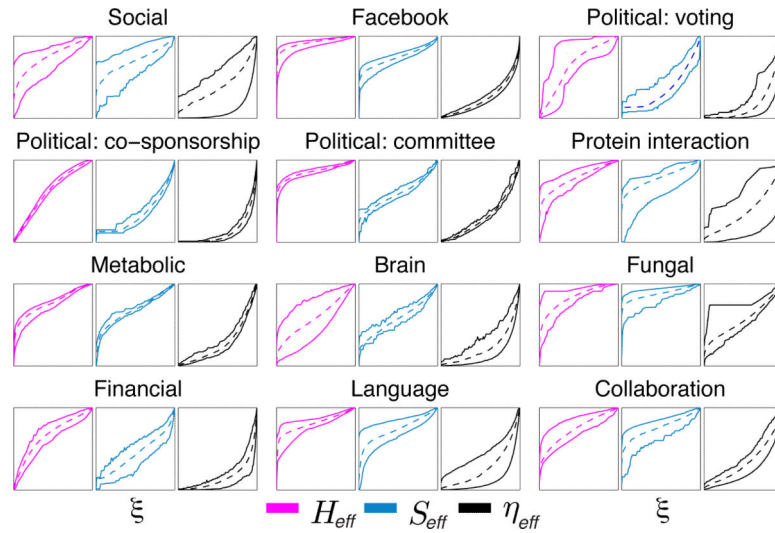


FIG. 7. (Color online) MRFs for all of the network categories containing at least 8 networks (see Table I). At each value of ξ , the upper curve shows the maximum value of \mathcal{H}_{eff} (pink, left panel in each category), S_{eff} (blue, center panel), and η_{eff} (black, right panel) for all networks in the category and the lower curve shows the minimum value. The dashed curves show the corresponding mean MRFs.

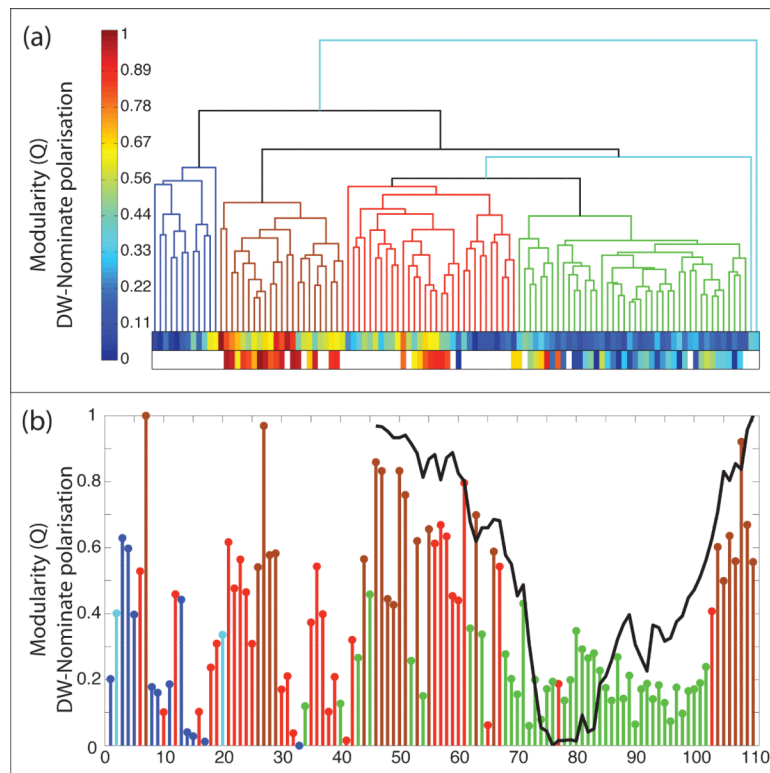


FIG. 8.

(Color online) (a) Dendrogram for Senate roll-call voting networks for the 1st–110th Congresses. Each leaf in the dendrogram represents a single Senate. Two horizontal color bars below the dendrograms indicate polarization measured in terms of modularity (upper bar) and DW-Nominate scores (lower bar). We color the branches in the dendrogram corresponding to periods of similar polarization. (b) Polarization of the US Senate as a function of time. The height of each stem indicates the level of polarization measured using modularity, and the color of each stem gives the cluster membership of each Senate in (a). The black curve shows the DW-Nominate polarization. Note that we have rescaled both measures to the interval $[0, 1]$.

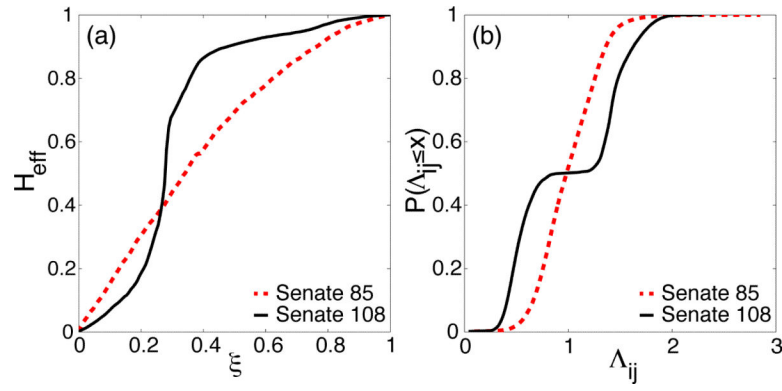


FIG. 9. (Color online) Comparison of the (low-polarization) 85th Senate and the (high-polarization) 108th Senate. The panels show (a) the \mathcal{H}_{eff} MRFs and (b) the cumulative distributions of Λ_{ij} values.

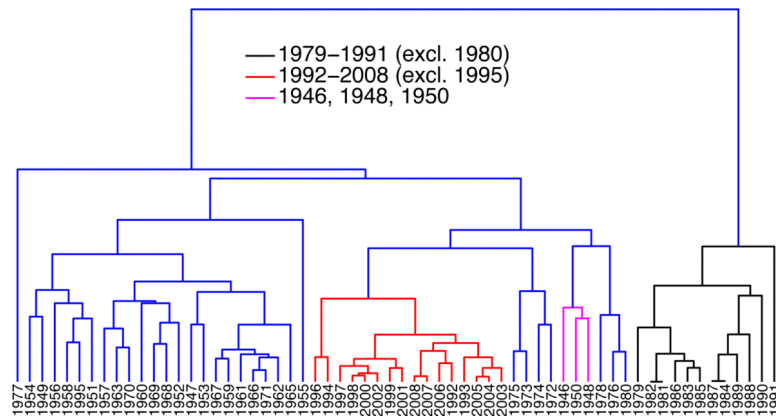


FIG. 10.
(Color online) Dendrogram for the United Nations General Assembly resolution voting network for the 1st–63rd sessions (excluding the 19th session), covering the period 1946–2008. Each leaf in the dendrogram represents a single session. In the text, we discuss the coloring of groups of branches in the dendrogram.

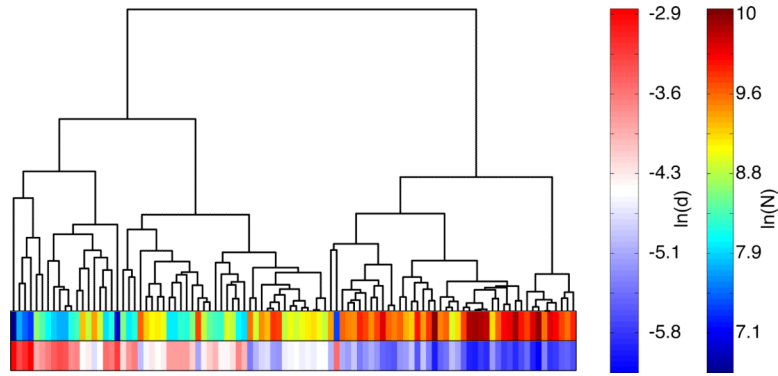


FIG. 11. (Color online) Dendrogram for 100 Facebook networks of US universities at a single-time snapshot in September 2005. We order the leaves of the dendrogram to minimize the distance between adjacent nodes. The color bars below the dendrogram indicate (top) the number of nodes in the networks N and (bottom) the fraction of possible edges that are present d .

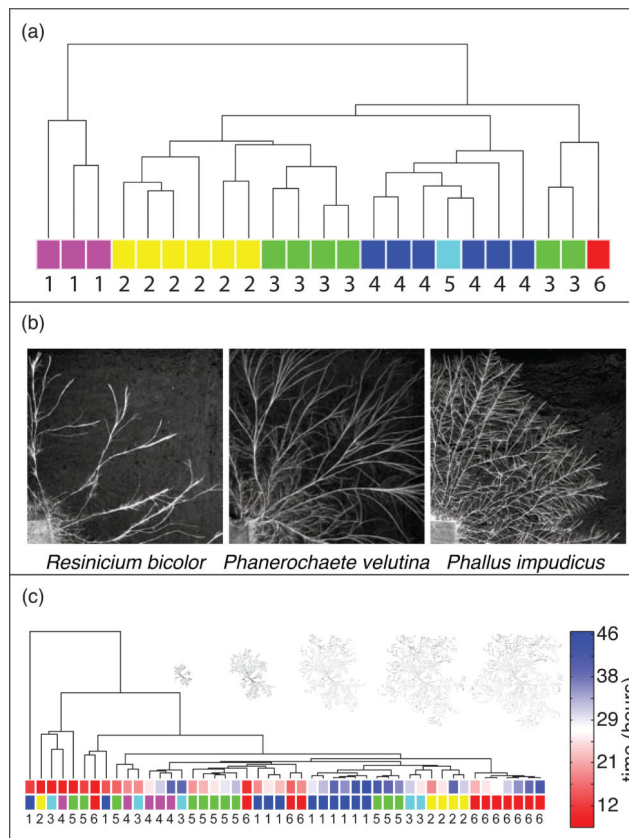


FIG. 12.

(Color online) (a) Dendrogram of networks for six different species of *Saprotrrophic basidiomycetes* and the slime mold *Physarum polycephalum*. Each leaf represents a replicate experiment. The colors and numbers correspond to the species as follows: (1) *Resinicium bicolor*, (2) *Physarum polycephalum*, (3) *Phallus impudicus*, (4) *Phanerochaete velutina*, (5) *Stropharia caerulea*, and (6) *Agrocybe gibberosa*. (b) Images illustrating the network structure of the different species [52]. (c) Dendrogram of network development in six replicate time series of *Phanerochaete velutina*. We color the leaves by time, and the color bar underneath the leaves indicates experiment number (1, . . . , 6). In the inset, we show extracted networks that illustrate the transition from simple branching trees to increasing levels of interconnection (i.e., cross-linking) with time.

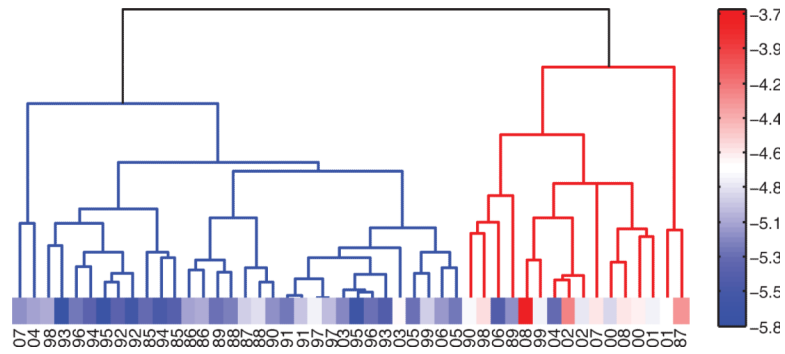


FIG. 13.
(Color online) Dendrogram for 48 NYSE networks during the period 1985–2008 [22].
Observe the clear split of the dendrogram into two clusters (a blue group on the left and a red group on the right). Leaf color indicates mean daily volatility of the composite index.

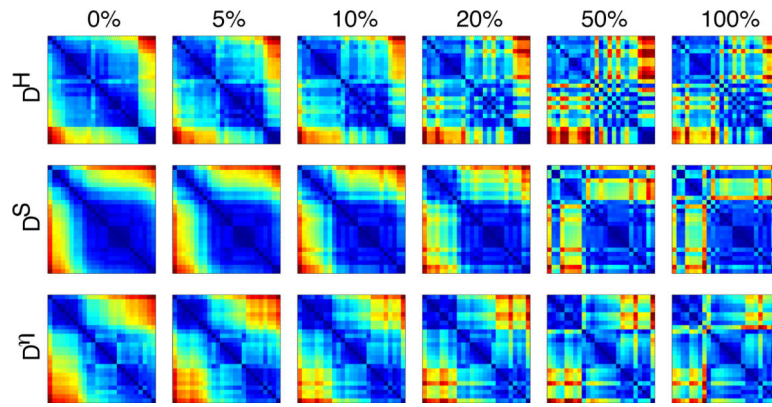


FIG. 14.

(Color online) Block-diagonalized mean distance matrices $\mathbf{D}^{\mathcal{H}}$ (top row), \mathbf{D}^S (middle row), and \mathbf{D}^U (bottom row) for the 25 networks listed in bold in Table II of the Supplemental Material. The columns show the mean-distance matrices following randomizations of the original network in which a given percentage of edges are rewired and the degree distributions of the networks are preserved. (We also constrain each rewired network to consist of a single connected component.) The ordering of the nodes in the plots is fixed for each row. The first column shows the distance matrix for the original networks. The distance matrices for the randomizations are the mean pairwise distances between networks.

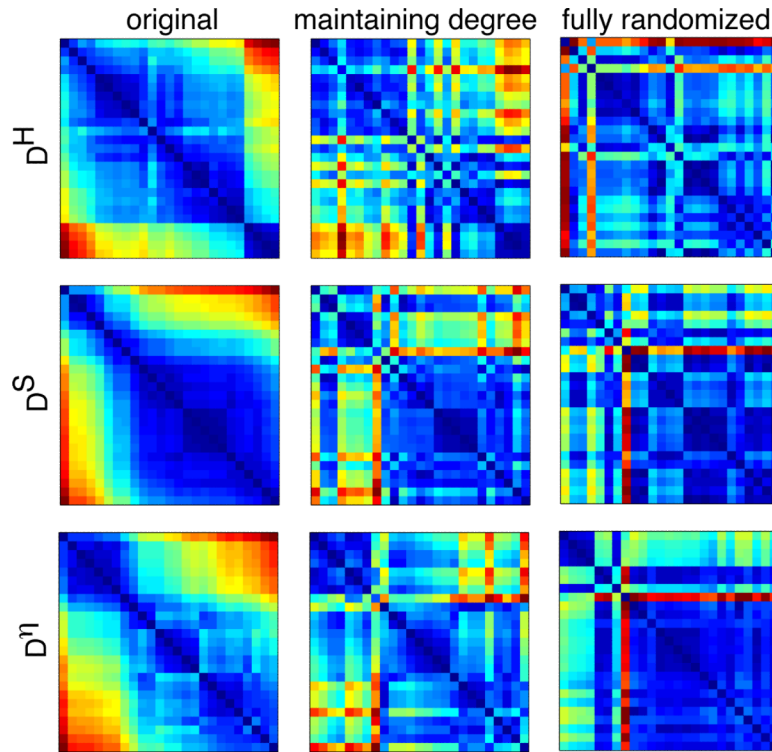


FIG. 15. (Color online) Block-diagonalized distance matrices $\mathbf{D}^{\mathcal{H}}$ (top row), \mathbf{D}^S (middle row), and \mathbf{D}^{η} (bottom row) for the 25 networks listed in bold in Table II of the Supplemental Material. The first column shows the distance matrices for the original networks. The second column shows the mean distance matrices following randomizations of the original networks in which 10 times the total number of edges in the networks have been rewired such that the degree distributions are preserved and the rewired networks each consist of a single connected component. The third column shows the mean distance matrices following randomizations of the original networks in which 10 times the total number of edges in the networks have been rewired but only the fact that the networks consist of single connected components is preserved (i.e., the degree distributions are not preserved). The distance matrices for the randomizations are composed of the mean pairwise distances between the networks.

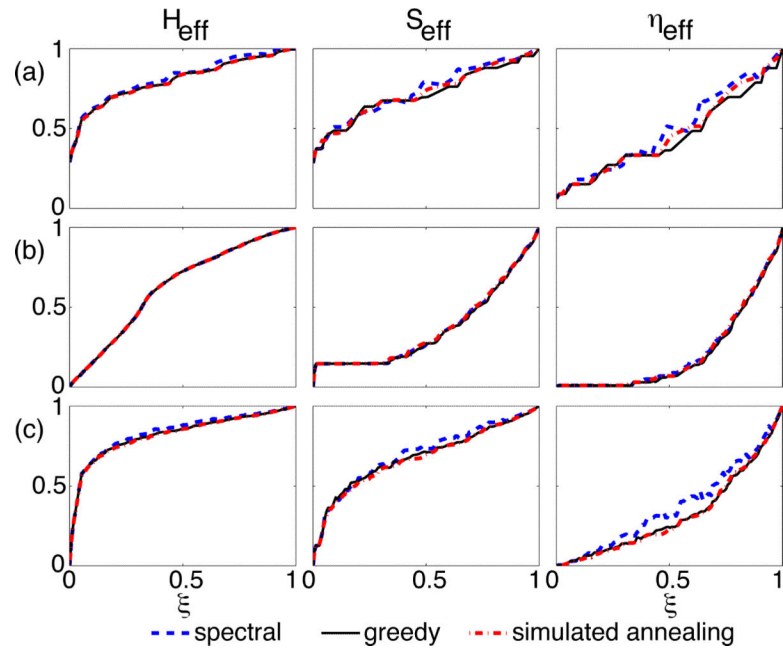


FIG. 16. (Color online) Comparison of the MRFs produced using spectral [18], Louvain [25], and simulated annealing [58] optimization heuristics. We show the MRFs for the (a) Zachary Karate Club network [83], (b) the roll-call voting network of the 102nd US Senate [50–52], and (c) the Garfield small-world citations network [40].

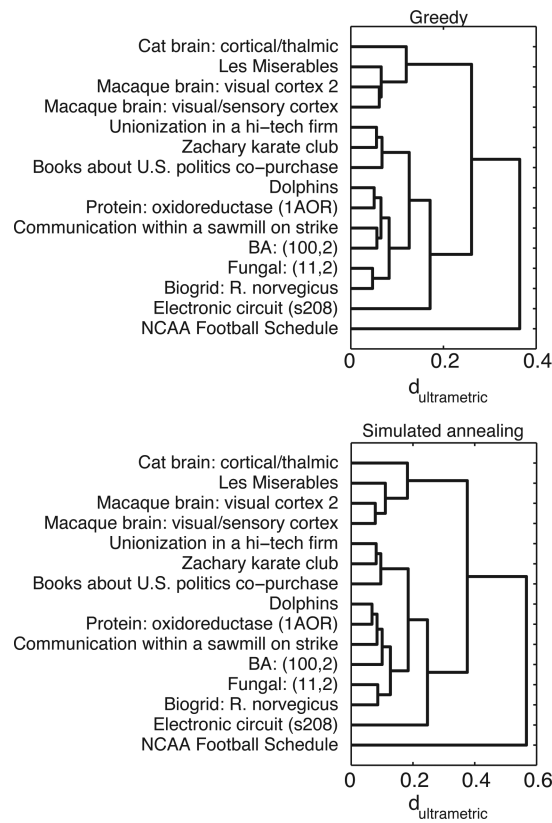


FIG. 17.

(Color online) Comparison of the dendrograms produced using a Louvain algorithm (top panel) and simulated annealing (bottom panel) for a subset of 15 networks. The only difference between the two dendrograms is the order in which the “Communication within a sawmill on strike” and the “BA: (100,2)” networks cluster and the distances at which the other networks cluster.

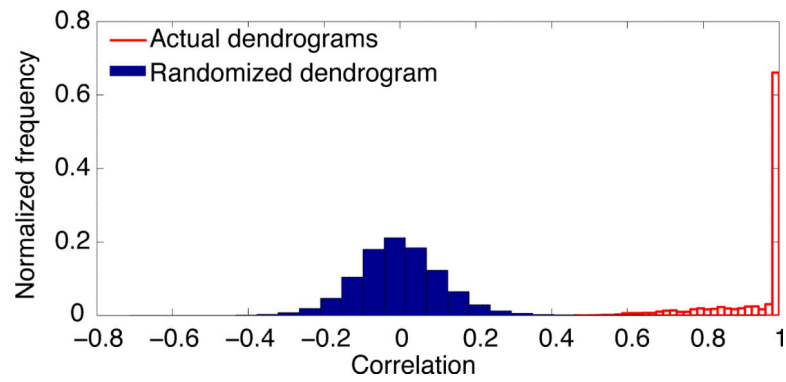


FIG. 18. (Color online) Comparison of the distributions of correlation coefficients between empirical Louvain dendrograms and empirical (red, hollow) and randomized (blue, solid) simulated-annealing dendrograms. See the text for details.

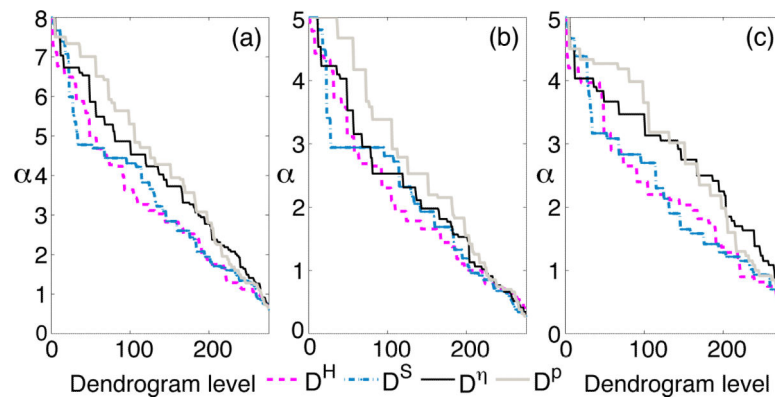


FIG. 19.

(Color online) Comparison of the effectiveness of the employed distance measures at clustering networks of the same category. As discussed in this text, we quantify this using the clustering diagnostic $\alpha(h)$. We calculate dendrograms from four distance matrices (\mathbf{D}^H , \mathbf{D}^S , \mathbf{D}^η , and \mathbf{D}^p) and compare the resulting values of $\alpha(h)$ for different sets of categories. **(a)** The value of the clustering diagnostic $\alpha(h)$ as a function of dendrogram cut level h (i.e., where the dendrogram is split to clusters) for the following 8 categories of networks: Facebook, metabolic, political cosponsorship, political committee, political voting, financial, brain, and fungal. **(b)** The value of $\alpha(h)$ for the largest 5 of the above 8 categories (Facebook, metabolic, political cosponsorship, political committee, and political voting) and **(c)** for the smallest 5 of the above 8 categories (Facebook, metabolic, financial, brain, and fungal). The maximum possible value of $\alpha(h)$ in each panel is equal to the number of categories considered in each panel. The values of $\alpha(h)$ obtained using the PCA-distance matrix \mathbf{D}^p (gray solid curve) are usually higher than those obtained using the other three distance measures. This suggests that PCA distance is the most effective of the four employed clustering measures.

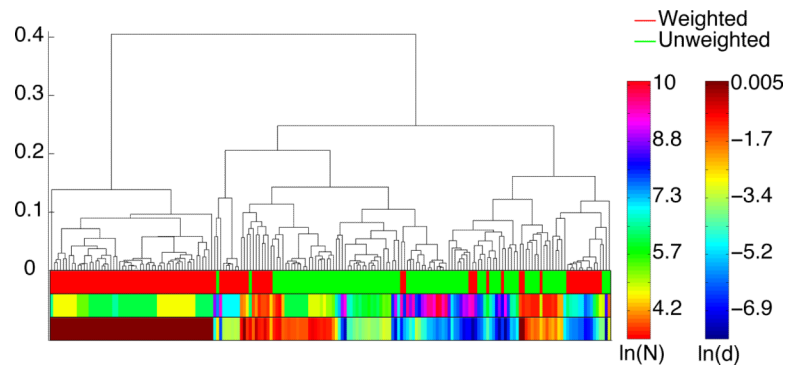


FIG. 20. (Color online) Taxonomy for 189 networks. We constructed the dendrogram using the distance matrix \mathbf{D}^p and average linkage clustering. We order the leaves of the dendrogram to minimize the distance between adjacent nodes, and we color the leaves to indicate the type of network. The three color bars below the dendrogram indicate whether the network corresponding to each leaf is weighted or unweighted (top), the number of nodes in the networks N (middle), and the fraction of possible edges that are present d (bottom).

TABLE I

Network categories, the total number of networks assigned to each category, and the number of networks from each category included in the taxonomy in Fig. 6. For the full taxonomy that uses all 746 networks, see Fig. 1 of the Supplemental Material.

Category	All networks	Taxonomy networks
Political: voting	285	23
Facebook	100	15
Fungal	65	12
Synthetic	58	0
Financial	54	6
Metabolic	43	15
Social	26	26
Political: cosponsorship	26	26
Other	23	0
Protein interaction	22	22
Political: committee	16	16
Brain	12	12
Language	8	8
Collaboration	8	8
Total	746	189